Pliny User Manual

Table of Contents

1. Introduction
2. Getting Started
3. Using the Application
4. Understanding the Visualizations
5. Advanced Features
6. Troubleshooting
7. FAQ
8. Glossary

1. Introduction

1.1 About Pliny

Pliny is a web-based application designed to visualize trends in Wikipedia metadata over time. Named after Pliny the Elder, a Roman author known for his encyclopedic work "Natural History," this tool provides insights into what people around the world are reading, editing, and contributing to on Wikipedia.

1.2 Key Features

- Daily-updated visualizations of Wikipedia trends
- Historical data access dating back to September 2023
- Multiple visualization types showing different aspects of Wikipedia activity
- User-friendly interface for exploring metadata patterns
- Category-based filtering of trends
- Date selection for historical analysis

1.3 Why Wikipedia Metadata Matters

Wikipedia is the fifth most visited website globally, serving as a primary source of information for millions of people daily. By analyzing metadata (information about page views, edits, and other activity), Pliny reveals patterns in global information consumption and knowledge creation that would otherwise remain hidden.

2. Getting Started

2.1 System Requirements

Pliny is a web-based application that works in any modern web browser. For optimal experience, we recommend: - Chrome 90+, Firefox 88+, Safari 14+, or Edge 90+ - JavaScript enabled - Screen resolution of 1280×720 or higher - Stable internet connection

2.2 Accessing Pliny

To access Pliny, visit pliny.wiki in your web browser. The application is freely available with no login required.

2.3 Interface Overview

The Pliny interface consists of several key elements:

1. **Header Bar**: Contains the application title and navigation
2. **Date Selector**: Allows you to choose which date's data to view
3. **Visualization Area**: The main content area displaying various trend visualizations
4. **Category Filter**: Optional filters to focus on specific topic areas
5. **Information Panels**: Contextual information about displayed trends

3. Using the Application

3.1 Selecting a Date

1. Locate the date selector at the top of the page
2. Click on the calendar icon to open the date picker
3. Select any date from September 1, 2023, to the present
4. The visualizations will automatically update to show data from your selected date

Note: Dates in the future or before September 2023 are not available for selection as data collection began in September 2023.

3.2 Filtering by Category

1. Locate the category dropdown in the header area
2. Click to open the category list
3. Select a category (e.g., Science, Entertainment, Politics) to filter the displayed trends
4. To view all categories again, select "All Categories"

3.3 Interacting with Visualizations

Most visualizations in Pliny are interactive. Common interactions include:

- **Hovering**: Displays detailed information about specific data points
- **Clicking**: Selects a specific trend for detailed view
- **Zooming**: Some charts allow zooming for detailed inspection
- **Panning**: After zooming, you can pan to see different parts of the visualization

3.4 Sharing Insights

To share a specific view or finding: 1. Navigate to the desired date and category 2. Copy the URL from your browser's address bar 3. Share this URL with others to show them the exact same view

4. Understanding the Visualizations

4.1 Trending Pages

This visualization shows Wikipedia pages with unusual spikes in viewership compared to their historical averages.

How to read it: - Each bar represents a Wikipedia page - The length of the bar indicates the magnitude of the viewership spike - Color coding may indicate different categories or topics - Hover over a bar to see the actual number of views and percentage increase

What it tells you: Pages with sudden viewership spikes often reflect current events, media releases, or viral topics capturing public attention.

4.2 Edit Activity

This visualization displays patterns in Wikipedia editing behavior, showing which pages are being actively modified.

How to read it: - Each point typically represents a Wikipedia page - The size of the point may indicate the volume of edits - Clustering indicates related pages being edited together - Color coding may represent different editor types or edit categories

What it tells you: High edit activity often indicates evolving topics, controversies, or new information being integrated into Wikipedia's knowledge base.

4.3 Category Distribution

This visualization shows how page views or edit activity distributes across different Wikipedia categories.

How to read it: - Segments represent different categories - The size of each segment indicates relative activity - Hover over segments for exact percentages

What it tells you: This reveals which knowledge domains are receiving the most attention from Wikipedia users during the selected time period.

4.4 Time Series Trends

This visualization shows how specific metrics change over time.

How to read it: - The x-axis represents time (typically days) - The y-axis represents the measured value (views, edits, etc.) - Multiple lines may represent

different pages or categories - Points along the line indicate specific measurements

What it tells you: Time series reveal patterns, cycles, and unusual events in Wikipedia usage over time.

5. Advanced Features

5.1 Comparing Time Periods

To compare trends across different dates: 1. Note the key metrics for your first date of interest 2. Use the date selector to choose a different date 3. Compare the visualizations between the two dates

Note: While Pliny doesn't currently support side-by-side comparison, this manual approach allows for basic comparison.

5.2 Understanding Statistical Significance

Pliny uses statistical methods to identify meaningful trends rather than random fluctuations. When a trend is highlighted, it represents a statistically significant deviation from expected patterns based on historical data.

The key methods used include: - Z-score analysis for pageview spikes - Clustering algorithms for related edit activity - Moving averages for baseline establishment

5.3 Using Pliny for Research

If you're using Pliny for academic or journalistic research, consider the following best practices: 1. Document the specific dates and filters you applied 2. Screenshot or save relevant visualizations 3. Cite Pliny as a data source in your references 4. Consider the limitations of the data (English Wikipedia only, data since 2023)

6. Troubleshooting

6.1 Visualizations Not Loading

If visualizations fail to load: 1. Check your internet connection 2. Try refreshing the page 3. Clear your browser cache 4. Try a different web browser 5. Ensure JavaScript is enabled

6.2 Date Selection Issues

If you cannot select a specific date: 1. Verify the date is within the available range (September 2023 to present) 2. Try selecting a nearby date first, then navigate to your target date 3. Refresh the page and try again

6.3 Visual Display Problems

If visualizations appear distorted: 1. Try resizing your browser window 2. Check your zoom level (100% is recommended) 3. Try landscape orientation on mobile devices 4. Disable browser extensions that might interfere with visualizations

7. FAQ

7.1 General Questions

Q: Is Pliny affiliated with Wikipedia or the Wikimedia Foundation?
A: No, Pliny is an independent project that analyzes publicly available Wikipedia metadata.

Q: Does Pliny cost anything to use?
A: No, Pliny is completely free to use.

Q: How often is the data updated?
A: Data is updated daily, typically with a 24-hour delay from real-time Wikipedia activity.

Q: Does Pliny work with non-English Wikipedia editions?
A: Currently, Pliny only analyzes English Wikipedia data. Support for other languages is planned for future releases.

7.2 Data Questions

Q: Where does the data come from?
A: All data comes from publicly available Wikipedia APIs and data dumps.

Q: How far back does the historical data go?
A: Currently, data is available from September 1, 2023, onward.

Q: Is the raw data available for download?
A: Pliny currently doesn't offer raw data downloads, but the visualizations reflect the processed data.

Q: How does Pliny determine what's "trending"?
A: Pliny uses statistical methods to identify unusual patterns in page views or edit activity compared to historical baselines.

7.3 Technical Questions

Q: Is Pliny open source?
A: Yes, the Pliny codebase is open source and available on GitHub.

Q: Can I contribute to Pliny?
A: Yes, contributions are welcome. Visit the GitHub repository for details on how to contribute.

Q: What technologies does Pliny use?
A: Pliny uses Python for data processing, Go for the backend, and React/TypeScript for the frontend. Data is stored in Google BigQuery.

8. Glossary

BigQuery: Google's enterprise data warehouse used by Pliny to store and analyze Wikipedia metadata.

Category: A classification of Wikipedia pages by subject matter (e.g., Science, Entertainment).

Edit: A change made to a Wikipedia page, which can range from minor corrections to major content additions.

Metadata: Data about Wikipedia activity, such as page views and edits, rather than the actual content of articles.

Page view: A single instance of a user accessing a Wikipedia page.

Spike: A sudden, significant increase in activity (usually page views) over a short period.

Trend: A pattern of change in Wikipedia metadata over time that reveals insights about user behavior.

Visualization: A graphical representation of data designed to make patterns and insights easily understandable.

Wikipedia API: Application Programming Interfaces provided by Wikipedia that allow access to their data.

Z-score: A statistical measurement used to identify how many standard deviations a data point is from the mean, used by Pliny to identify unusual activity.